Simulation-to-Experiment Transfer Learning for XRD-Based Autoencoder Modeling of SOEC Half-Cell Materials

Hunmin Park[1]

[1]*Korea Institute of Energy Research, 152 Gajeong-ro, Yuseong-gu, Daejeon (34129)*

Correspondence to: Hunmin Park hmpark@kier.re.kr

1. Introduction

The development of materials for solid oxide electrolysis cells (SOECs) is hindered by several challenges, including the difficulty of non-destructive quality inspection, limited capability for performance prediction, high costs and long development times associated with high-temperature and multi-step fabrication processes, and the absence of dedicated, high-reliability artificial intelligence (AI) models tailored for SOEC materials. To overcome these limitations, this study proposes a data-driven materials AI framework based on an autoencoder architecture for X-ray diffraction (XRD) analysis, aiming to enable non-destructive evaluation and to establish a foundation for future process identification and performance prediction.

2. Experimental

An autoencoder model was designed to learn latent representations of XRD patterns, where the encoder extracts essential crystallographic features and compresses them into a laten space, and the decoder reconstructs the original XRD patterns from these features. For dataset construction, 50 SOEC half-cells were fabricated and their experimental XRD patterns were collected. In addition, 10,000 simulated XRD patterns with similar crystallographic characteristics were generated and used to pre-train the autoencoder. The pre-trained model was subsequently fine-tuned using the experimental XRD data to improve its adaptability to real experimental conditions.

3. Result and Discussion

3.1 Autoencoder Reconstruction Performance

The autoencoder model pre-trained on simulated XRD data demonstrated stable reconstruction performance, indicating that the encoder successfully extracted essential crystallographic features from the input patterns. The reconstructed XRD patterns closely matched the original simulated inputs, confirming that the latent space effectively preserved peak position, relative intensities, and overall diffraction profiles. This result suggests that the autoencoder architecture is suitable for learning compact yet informative representations of SOEC-related XRD data.

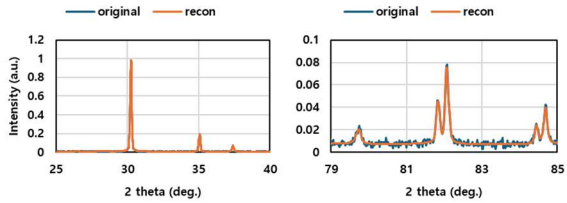

Fig. 1: Comparison between input XRD patterns and reconstructed XRD patterns generated by the autoencoder pre-trained on simulated XRD data. The reconstructed patterns (orange lines) closely match the input patterns (blue lines) in both peak positions and intensities, demonstrating effective feature learning and reconstruction capability of the pre-trained model.

3.2 Similarity Between Simulated and Experimental XRD patterns

When the pre-trained model was applied to experimental XRD data, the reconstructed patterns showed a high degree of similarity to the measured experimental patterns. Notably, the model generated XRD patterns that not only reproduced the dominant diffraction peaks but also captured subtle variations in peak intensity and shape. These results indicate that simulation-based pre-training can provide physically meaningful prior knowledge, enabling the model to bridge the gap between idealized simulated data and experimentally obtained XRD patterns.

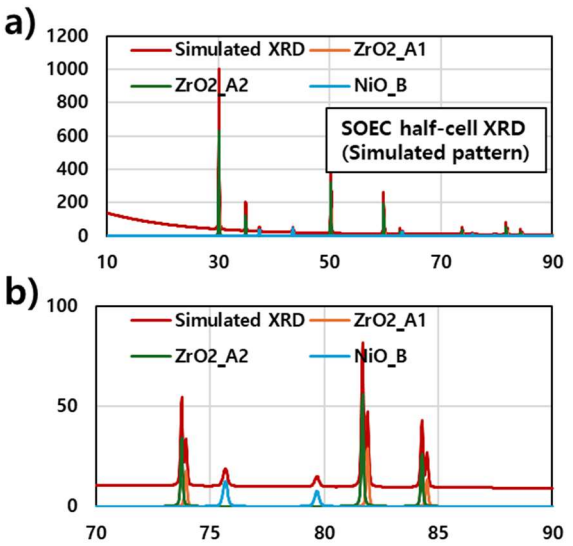

Fig. 2: (a) Comparison between simulated XRD and representative experimental XRD patterns obtained from SOEC half-cells, demonstrating overall agreement in peak positions and relative intensities. (b) Decomposition of the simulated XRD pattern into constituent phases, including ZrO_2 and NiO components, illustrating the phase-wise contribution to the overall diffraction profile

3.3 Effect of Fine-Tuning with Experimental XRD data

To further improve model fidelity, the pre-trained autoencoder was fine-tuned using experimental XRD data obtained from 50 fabricated SOEC half-cells. After fine-tuning, the reconstruction error was significantly reduced compared to the pre-training stage. The fine-tuned model exhibited enhanced sensitivity to experimental noise, minor peak shifts, and intensity fluctuations, which are commonly observed in real measurements but are difficult to fully represent in simulated data. This improvement demonstrates that fine-tuning effectively adapts the latent space to experimental conditions while retaining the generalized learned during pre-training.

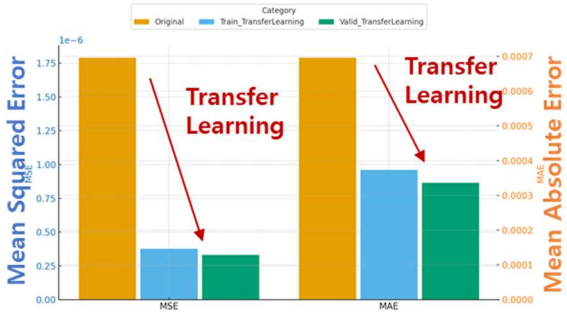

Fig. 3: Performance comparison demonstrating the effect of transfer learning on autoencoder reconstruction accuracy. The pretrained model (orange bars) is compared with the fine tuned model evaluated on training data (light blue bars) and validation data (green bars).

3.4 Implications of Latent Space Representations

The learned latent space can be interpreted as a compressed representation of the structural characteristics of SOEC half-cell materials. Because the latent features are derived from XRD patterns, they inherently encode information related to phase composition, crystallinity, and microstructural variations. This representation provides a promising basis for downstream AI models aimed at identifying fabrication or processing conditions and predicting electrochemical performance without requiring destructive testing or additional high-cost experiments.

3.5 Outlook for Process Identification and Performance Prediction

The proposed XRD-based autoencoder framework establishes a scalable pathway for integrating materials AI into SOEC development. By leveraging simulation-to-experiment transfer learning, the approach reduces the reliance on large experimental datasets while maintaining high reconstruction accuracy. In future work, the latent space features extracted from the trained autoencoder can be coupled with supervised learning models to enable automated process condition classification and quantitative performance prediction of SOEC half-cells. This approach is expected to significantly reduce development time and cost, accelerating the optimization of SOEC materials and manufacturing processes.